Class Assignment 4

Arturo Ortiz

2023-11-01

crime\_ds = read.csv("https://raw.githubusercontent.com/EricBrownTTU/ISQS6350/main/crime.csv")  
crime\_ds

## STATE MURDER RAPE ROBBERY ASSAULT BURGLARY LARCENY AUTO  
## 1 ALABAMA 14.2 25.2 96.8 278.3 1135.5 1881.9 280.7  
## 2 ALASKA 10.8 51.6 96.8 284.0 1331.7 3369.8 753.3  
## 3 ARIZONA 9.5 34.2 138.2 312.3 2346.1 4467.4 439.5  
## 4 ARKANSAS 8.8 27.6 83.2 203.4 972.6 1862.1 183.4  
## 5 CALIFORNIA 11.5 49.4 287.0 358.0 2139.4 3499.8 663.5  
## 6 COLORADO 6.3 42.0 170.7 292.9 1935.2 3903.2 477.1  
## 7 CONNECTICUT 4.2 16.8 129.5 131.8 1346.0 2620.7 593.2  
## 8 DELAWARE 6.0 24.9 157.0 194.2 1682.6 3678.4 467.0  
## 9 FLORIDA 10.2 39.6 187.9 449.1 1859.9 3840.5 351.4  
## 10 GEORGIA 11.7 31.1 140.5 256.5 1351.1 2170.2 297.9  
## 11 HAWAII 7.2 25.5 128.0 64.1 1911.5 3920.4 489.4  
## 12 IDAHO 5.5 19.4 39.6 172.5 1050.8 2599.6 237.6  
## 13 ILLINOIS 9.9 21.8 211.3 209.0 1085.0 2828.5 528.6  
## 14 INDIANA 7.4 26.5 123.2 153.5 1086.2 2498.7 377.4  
## 15 IOWA 2.3 10.6 41.2 89.8 812.5 2685.1 219.9  
## 16 KANSAS 6.6 22.0 100.7 180.5 1270.4 2739.3 244.3  
## 17 KENTUCKY 10.1 19.1 81.1 123.3 872.2 1662.1 245.4  
## 18 LOUISIANA 15.5 30.9 142.9 335.5 1165.5 2469.9 337.7  
## 19 MAINE 2.4 13.5 38.7 170.0 1253.1 2350.7 246.9  
## 20 MARYLAND 8.0 34.8 292.1 358.9 1400.0 3177.7 428.5  
## 21 MASSACHUSETTS 3.1 20.8 169.1 231.6 1532.2 2311.3 1140.1  
## 22 MICHIGAN 9.3 38.9 261.9 274.6 1522.7 3159.0 545.5  
## 23 MINNESOTA 2.7 19.5 85.9 85.8 1134.7 2559.3 343.1  
## 24 MISSISSIPPI 14.3 19.6 65.7 189.1 915.6 1239.9 144.4  
## 25 MISSOURI 9.6 28.3 189.0 233.5 1318.3 2424.2 378.4  
## 26 MONTANA 5.4 16.7 39.2 156.8 804.9 2773.2 309.2  
## 27 NEBRASKA 3.9 18.1 64.7 112.7 760.0 2316.1 249.1  
## 28 NEVADA 15.8 49.1 323.1 355.0 2453.1 4212.6 559.2  
## 29 NEW HAMPSHIRE 3.2 10.7 23.2 76.0 1041.7 2343.9 293.4  
## 30 NEW JERSEY 5.6 21.0 180.4 185.1 1435.8 2774.5 511.5  
## 31 NEW MEXICO 8.8 39.1 109.6 343.4 1418.7 3008.6 259.5  
## 32 NEW YORK 10.7 29.4 472.6 319.1 1728.0 2782.0 745.8  
## 33 NORTH CAROLINA 10.6 17.0 61.3 318.3 1154.1 2037.8 192.1  
## 34 NORTH DAKOTA 0.9 9.0 13.3 43.8 446.1 1843.0 144.7  
## 35 OHIO 7.8 27.3 190.5 181.1 1216.0 2696.8 400.4  
## 36 OKLAHOMA 8.6 29.2 73.8 205.0 1288.2 2228.1 326.8  
## 37 OREGON 4.9 39.9 124.1 286.9 1636.4 3506.1 388.9  
## 38 PENNSYLVANIA 5.6 19.0 130.3 128.0 877.5 1624.1 333.2  
## 39 RHODE ISLAND 3.6 10.5 86.5 201.0 1489.5 2844.1 791.4  
## 40 SOUTH CAROLINA 11.9 33.0 105.9 485.3 1613.6 2342.4 245.1  
## 41 SOUTH DAKOTA 2.0 13.5 17.9 155.7 570.5 1704.4 147.5  
## 42 TENNESSEE 10.1 29.7 145.8 203.9 1259.7 1776.5 314.0  
## 43 TEXAS 13.3 33.8 152.4 208.2 1603.1 2988.7 397.6  
## 44 UTAH 3.5 20.3 68.8 147.3 1171.6 3004.6 334.5  
## 45 VERMONT 1.4 15.9 30.8 101.2 1348.2 2201.0 265.2  
## 46 VIRGINIA 9.0 23.3 92.1 165.7 986.2 2521.2 226.7  
## 47 WASHINGTON 4.3 39.6 106.2 224.8 1605.6 3386.9 360.3  
## 48 WEST VIRGINIA 6.0 13.2 42.2 90.9 597.4 1341.7 163.3  
## 49 WISCONSIN 2.8 12.9 52.2 63.7 846.9 2614.2 220.7  
## 50 WYOMING 5.4 21.9 39.7 173.9 811.6 2772.2 282.0

# Problem #1

The first six components account for about 98.2% of the variability. So, the use of these six components to capture a major percentage of the variability in the data would be reasonable.

numeric\_data <- crime\_ds[, sapply(crime\_ds, is.numeric)]  
  
numeric\_data\_std <- scale(numeric\_data)  
  
pca\_results <- princomp(numeric\_data\_std)  
  
cumulative\_variance <- cumsum(pca\_results$sdev^2) / sum(pca\_results$sdev^2)  
  
print(cumulative\_variance)

## Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7   
## 0.5878514 0.7648116 0.8684997 0.9137043 0.9505578 0.9822777 1.0000000

# Problem #2

Component 2 contributes considerably to the dataset’s variability, specifically in terms of violent crimes, making it a more impactful element than Component 1.

# Problem #3

The biplot shows that Hawaii (11) and Mississippi (24) are at opposite ends of the plot vertically. Hawaii is at the bottom of the chart, while Mississippi is at the top. For these two situations, the direction of the arrows indicating the variables is completely opposite. This implies that the values of the variables vary significantly between Hawaii and Mississippi. In particular, the factors represented by the arrows are significantly lower in Hawaii than in Mississippi. Nevada (28), is located as far right as possible on the biplot. This fits with the direction of the arrows for the variables, suggesting that Nevada has much higher values for these specific variables. It is located directly between the dots referencing rape and robbery, meaning that Nevada has far higher rates of both crimes than Hawaii and Mississippi.

biplot(pca\_results, col = c("blue", "red"), cex = 0.6)

